# Beyond cyanogenesis: Temperature gradients drive environmental adaptation in North American white clover (*Trifolium repens* L.)

Running title (<= 45 characters): Landscape genomics of white clover

Wen-Hsi Kuo (0000-0003-4680-1105)1\*

Limei Zhong (0000-0002-7472-5245)2\*

Sara J. Wright (0000-0001-5864-2661)1,3

David M. Goad (0000-0001-8658-6660)1

Kenneth M. Olsen (0000-0002-8338-3638)1\*\*

1Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA

2Key Laboratory of Molecular Biology and Gene Engineering in Jiangxi, School of Life Sciences, Nanchang University, Nanchang, China

3Present address: Department of Biological and Biomedical Sciences, Rowan University, Glassboro, New Jersey, USA

\*Wen-Hsi Kuo and Limei Zhong contributed equally to this work

\*\*Corresponding author

[wen-hsi.kuo@wustl.edu](mailto:wen-hsi.kuo@wustl.edu)

[limei\_zhong@qq.com](mailto:limei_zhong@qq.com)

[wrights@rowan.edu](mailto:wrights@rowan.edu)

[dgoad90@gmail.com](mailto:dgoad90@gmail.com)

[kolsen@wustl.edu](mailto:kolsen@wustl.edu)

## Funding information

W-HK was funded through the William H. Danforth Plant Science Graduate Research Fellowship in the Division of Biology and Biomedical Sciences at Washington University, and by a scholarship from the Taiwan Ministry of Education.  Additional funding for the project was provided by US National Science Foundation grants IOS-1557770 to KMO, and DEB-1601641 and DGE-1143954 to SJW.

## Conflict of interest statement

The authors declare no conflict of interest.

# Abstract

Species that repeatedly evolve phenotypic clines across environmental gradients have been highlighted as ideal systems for characterizing the genomic basis of local environmental adaptation. However, few studies have assessed the importance of observed phenotypic clines for local adaptation: conspicuous traits that vary clinally may not necessarily be the most critical in determining local fitness. The present study was designed to fill this gap, using a plant species characterized by repeatedly-evolved adaptive phenotypic clines. White clover is naturally polymorphic for its chemical defense cyanogenesis (HCN release with tissue damage); climate-associated cyanogenesis clines have evolved throughout its native and introduced range worldwide. We performed landscape genomic analyses on 415 wild genotypes from 43 locations spanning much of the North American species range to assess the relative importance of cyanogenesis loci vs. other genomic factors in local climatic adaptation. We find clear evidence of local adaptation, with temperature-related climatic variables best describing genome-wide differentiation between sampling locations. The same climatic variables are also strongly correlated with cyanogenesis frequencies and gene copy number variations (CNVs) at cyanogenesis loci. However, landscape genomic analyses indicate no significant contribution of cyanogenesis loci to local adaptation. Instead, several genomic regions containing promising candidate genes for plant response to seasonal cues are identified — some of which are shared with previously-identified QTLs for locally-adaptive fitness traits in North American white clover. Our findings suggest that local adaptation in white clover is likely determined primarily by genes controlling the timing of growth and flowering in response to local seasonal cues. More generally, this work suggests that caution is warranted when considering the importance of conspicuous phenotypic clines as primary determinants of local adaptation.

# Keywords Local adaptation; cline; cyanogenesis; genotype-environment association (GEA), landscape genomics; temperature; white clover (*Trifolium repens*)

# Introduction

Understanding the phenotypic and genetic bases of local adaptation, where populations in their native habitats exhibit higher fitness compared to foreign populations, has emerged as a major focus in evolutionary biology (Kawecki & Ebert, 2004; Savolainen *et al.*, 2013; Lascoux *et al.*, 2016; Lasky *et al.*, 2022). One common approach is to employ reciprocal transplants and/or common garden experiments to evaluate fitness trade-offs between individuals originating from different environments (Stinchcombe *et al.*, 2004; Ågren *et al.*, 2013; Wright *et al.*, 2018; Wright *et al.*, 2022). When combined with genotypic information, such as quantitative trait locus (QTL) mapping of fitness traits, this approach can elucidate the genetic architecture of locally adaptive phenotypes (Ågren *et al.*, 2013; Grillo *et al.*, 2013; Wright *et al.*, 2022). A complementary approach for studying local adaptation is to employ landscape genomic methods to identify genomic regions with signatures of local adaptation; these may include loci that show evidence of adaptive differentiation between environments (e.g., FST outliers) or strong environmental correlations as detected in genotype-environment association (GEA) analyses (Guerrero *et al.*, 2018; Gugger *et al.*, 2018; reviewed in Lasky *et al.*, 2022; Battlay *et al.*, 2023; Wang *et al.*, 2023).

Among the different study systems that can potentially be used to study local adaptation, species that have repeatedly evolved phenotypic clines across an environmental gradient in different parts of their range have been held up as a “best-case scenario” for successfully detecting locally-adapted traits and their underlying genetic basis (Lasky *et al.*, 2022). In such species, the clinal variation is more likely to be reflecting a genuine signal of environmental adaptation than to be an artifact of neutral population structure or demographic history. However, even for such “best-case” study systems, it is important to recognize the distinction between traits that show clinal variation and traits that are most crucial for local adaptation: just because natural selection repeatedly favors the evolution of a phenotypic cline in a species, it does not necessarily follow that the obvious clinally-varying phenotype is most critical for determining fitness in local environments.

Nonetheless, very few studies have explicitly assessed the relative importance of conspicuous clincally-varying phenotypes for local adaptation. In species with repeatedly-evolved phenotypic clines, research has tended to focus on measuring the fitness effects of the known phenotypes, without knowledge of fitness contributions from other traits or loci (e.g., Mullen & Hoekstra, 2008; Linnen *et al.*, 2009; Antoniazza *et al.*, 2010; Wittkopp *et al.*, 2011; Hof *et al.*, 2016; Koski & Galloway, 2018; Gefaell *et al.*, 2023). Conversely, studies using landscape genomic methods to study local adaptation have generally not considered adaptive phenotypic clines (e.g., Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011; Kujala *et al.*, 2017; Price *et al.*, 2018; Battlay *et al.*, 2023). To address this gap, the present study applies landscape genomic analyses to study the genetic determinants of local adaptation in a geographically widespread plant species that features repeatedly-evolved phenotypic clines. By incorporating previously-collected data on fitness traits and QTLs for local environmental adaptation (Wright *et al.*, 2018; Wright *et al.*, 2022), we integrate insights from adaptive phenotypic clines, locally-adaptive fitness QTLs, and genomic signatures of local adaptation to holistically assess the genomic and phenotypic contributions to local environmental adaptation.

Our focal species, white clover (*Trifolium repens*, 2n = 4x = 32), is a geographically widespread herbaceous perennial legume that can be found in mesic regions worldwide. A native species of southern Europe, it was introduced globally as a forage and rotation crop and as a source of soil nitrogen, and it has since become widely naturalized as a component of lawns, roadsides and other mowed or grazed areas (Zeven, 1991; Kjærgaard, 2003). In both its native and introduced range, white clover populations can be found in abundance across a wide climatic range, spanning subtropical to boreal ecozones; this wide environmental adaptability has been attributed to white clover’s allopolyploid origin from two ecologically distinct diploid progenitor species (Griffiths *et al.*, 2019).

One clear indicator of white clover’s adaptive ability is that populations worldwide have evolved climate-associated clines in the chemical defense trait cyanogenesis (HCN release following tissue damage). The species is characterized by a genetically determined polymorphism for this defense (described below), with higher frequencies of cyanogenic plants generally present in warmer locations. Latitudinal and elevational cyanogenesis clines have been documented worldwide in numerous studies since the 1950s (Daday, 1954b; Daday, 1954a; Daday, 1958; Hughes, 1991; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013; Kuo *et al.*, 2024); more recently, urban-to-rural cyanogenesis clines have also been described (Thompson *et al.*, 2016; Santangelo *et al.*, 2022). While the selective factors that favor the rapid evolution of cyanogenesis clines have not been definitively determined, they likely involve fitness trade-offs in areas of high and low herbivore abundance, energetic costs associated with production of the required chemical precursors, and abiotic stress factors that could include freezing and drought (reviewed by Olsen *et al.*, 2013; Kooyers *et al.*, 2014; Kooyers *et al.*, 2018). Regardless of the specific selective factors at play, the fact that climate-associated cyanogenesis clines have evolved repeatedly throughout the species range suggests that selection on this polymorphism is strong, and that this trait could be an important determinant of local adaptation.

Despite the apparent strong selection favoring the repeated evolution of climate-associated cyanogenesis clines, our previous field experiments have not implicated the cyanogenesis polymorphism or its underlying genes as significant contributors to local adaptation (Wright *et al.*, 2018; Wright *et al.*, 2022). In a common garden experiment using 161 wild North American genotypes sampled from 15 locations spanning the North American species range, vegetative and reproductive fitness in a central US location were found to be strongly positively correlated with the climatic similarity of a genotype’s location-of-origin and the common garden site; however, the cyanogenesis polymorphism was only a negligible predictor of fitness in the common garden (Wright et al. 2018). In a second study, QTL mapping of fitness traits was performed using clonally-replicated F2 progeny of two biparental crosses between plants originating from the northern, central and southern US, with fitness measured in reciprocal common gardens in the locations of both parental genotypes. While genetic mapping of fitness traits revealed clear evidence of local adaptation and allelic trade-offs between environments, there was again little evidence that cyanogenesis variation contributes significantly to local adaptation (Wright *et al.*, 2022). However, both of these studies were necessarily limited by experimental design (see reviews by Savolainen *et al.*, 2013; Germino *et al.*, 2019) — most notably, a single common garden site in the study of Wright *et al.* (2018), and just three parental genotypes in the QTL mapping study of Wright et al. (2022). This leaves the possibility open that the lack of evidence for cyanogenesis polymorphism as a contributor to local adaptation is an artifact of limited resolution in the studies conducted to date.

To complement the previous common garden and QTL mapping approaches in white clover, the present study leveraged white clover’s recently published high-quality genome assembly (Kuo *et al.*, 2024) to conduct genotype environment association (GEA) analysis and genomic differentiation scans to test for signatures of local adaptation and the extent to which they do or do not involve cyanogenesis loci. Using a geographically representative sample of white clover genotypes from across North America (415 accessions from 43 locations), we sought to address four key questions: **1)** Is genome-wide genetic differentiation across the sampled range better explained by isolation-by-distance (IBD), which would be consistent with neutral population structure, or by isolation-by-environment (IBE), which would implicate environmental selection as a factor shaping the distribution of genotypes (Wang & Bradburd, 2014)? **2)** Does cyanogenesis clinal variation across the sampled range parallel genome-wide patterns of population differentiation, or are locus-specific patterns apparent for the genes controlling this chemical defense polymorphism? As a corollary, which environmental variable(s) can best predict cyanogenesis frequencies across the sampled populations? **3)** Do GEA and genomic differentiation scans identify the cyanogenesis loci as genetic contributors to local environmental adaptation? **4)** To what extent do genomic regions detected by the GEA and genomic differentiation scans overlap with fitness QTLs for local adaptation identified in our previous study (Wright *et al.*, 2022)? Are there candidate genes of known functions near the detected signals? To our knowledge, this study marks one of the first application of landscape genomic approaches to study local adaptation in a wild species characterized by adaptive phenotypic clines, allowing a comprehensive assessment of the relative importance of the clinal variation for local adaptation.

# Materials and Methods

## Sample Collection, DNA Extraction, and GBS Library Preparation

Using a nationwide network of K-12 science teachers, citizen scientists and colleagues, we obtained mature seeds or stolon cuttings for 419 wild white clover accessions across 43 locations in North America during the growing seasons of 2014-2017 (**Fig. 1**; **Table S1**). Each location was represented by 6 to 11 accessions (individual genotypes), and latitude and longitude were recorded for each sample. Seeds and stolon cuttings were cultivated in the greenhouse at Washington University in St. Louis under standard greenhouse conditions (see Wright *et al*. 2022). Genomic DNA was extracted from young leaves using a standard DNA extraction protocol (Whitlock *et al.*, 2008). Extraction quantity and quality were assessed using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, and Qubit™ dsDNA HS Assay Kits.

Genotyping-by-sequencing (GBS) libraries were prepared following Elshire *et al.* (2011) with *ApeK*I methylation sensitive restriction enzyme. Barcoding and protocol modifications are described in Olsen *et al.* (2021). Paired-end sequencing (150-bp reads) was performed using the Illumina Hi-Seq 2500 platform (Novogene Corp., Chula Vista, CA, USA).

## Read Mapping, SNP Calling, and SNP Filtering

Raw GBS reads were demultiplexed with SABRE (https://github.com/najoshi/sabre) and adaptor-trimmed with CUTADAPT (Martin, 2011). The processed reads were mapped back to our white clover reference genome (Kuo *et al.*, 2024) using BWA (Li & Durbin, 2009) with default paired-end settings. SNP calling from the alignments followed GATK best practices, with omission of the duplicated-read removal step as recommended for GBS data (Poplin *et al.*, 2017). The output SNP dataset in vcf format underwent hard filtering (bcftools filter -e 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || INFO/DP < 2500'). It was then filtered for sites with missing accessions < 0.25, missing sites < 0.35, and minor allele frequency > 0.05. Missing genotypes were imputed with Beagle v5.4 (Browning *et al.*, 2018). A relaxed Hardy-Weinberg filter was applied to remove sites with skewed heterozygosity suggesting sequencing error (p < 1×10-50). This dataset was used for GWAS and genome-wide environmental association (GEA). For population structure analyses and genomic differentiation scans, three different levels of linkage disequilibrium (LD) pruning were performed using PLINK2 (--indep-pairwise 100kb 0.8; --indep-pairwise 200kb 0.5; --indep-pairwise 500kb 0.2) before analysis (Chang *et al.*, 2015). Intrachromosomal LDs of pairwise SNPs in the range between >100 bp and <1000 kbp were calculated in PLINK1.9 (--r2 gz --ld-window 100 --ld-window-kb 1000 --ld-window-r2 0). Interchromosomal LDs were calculated by PLINK1.9 (--maf 0.2 --r2 inter-chr d --ld-window-r2 0).

## Population structure analysis and its association with local environmental variables

Population structure was analyzed based on the LD-pruned datasets. Ancestry estimation was performed by ADMIXTURE v1.3.0 with K = 1-8 and cross-validation mode (Alexander *et al.*, 2009). Principal component analysis (PCA) and pairwise FST estimation (between the 43 sampling locations) were conducted by PLINK2 (Chang *et al.*, 2015). To assess distributions of total genetic variation within and across population locations, an analysis of molecular variance (AMOVA) and its subsequent randomization test of significance (permutation = 1000) were carried out with the ‘poppr’ package in R (Kamvar *et al.*, 2014).

To explore the correlations between genetic differentiation, geographical distance, and environmental distance, a Mantel test and partial Mantel test (10,000 permutations apiece) were conducted with the ‘Vegan’ package in R (Dixon, 2003). Prior to the tests, FST was linearized as FST/(1-FST). Geographical distance was calculated as the great circle distance between sampling locations (using the mean value of geographical coordinates for all accessions in a given location) with the ‘sf’ package in R (Pebesma, 2018). The environmental distance was calculated as the Euclidean distance between the mean values of pairwise locations. Environmental values of sampling locations were derived from the 1000-m-buffered mean of WorldClim2 (Fick & Hijmans, 2017) and ENVIREM (Title & Bemmels, 2018) databases.

## Cyanogenesis polymorphism and its association with local environmental variables

The white clover cyanogenesis polymorphism is controlled by two independently-segregating simple Mendelian polymorphisms that control the presence/absence of two cyanogenic precursors, both of which must be present for a plant to be cyanogenic: *Ac/ac* controls the presence/absence of cyanogenic glucosides (stored in the vacuoles of photosynthetic tissue); and *Li/li* controls the presence/absence of their hydrolyzing enzyme, linamarase (present in the apoplast). In cyanogenic plants, tissue damage that causes cell rupture brings the two precursors together, leading to the liberation of HCN. At the molecular level, both the *Ac/ac* and *Li/li* polymorphisms correspond to gene presence/absence variation (PAVs), where recessive alleles correspond to genomic deletions (Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen & Small, 2018). Additionally, the dominant (functional) *Ac* and *Li* alleles both show tandem gene copy number variations (CNVs), which contribute to quantitative variation in the cyanogenic response (Kuo *et al.*, 2024). In wild North American populations, the CNVs at both loci show clinal variation in patterns that parallel climate-associated clines for the *Ac/ac* and *Li/li* PAVs (Kuo *et al.*, 2024).

Phenotyping and genotyping of the cyanogenesis polymorphism for the sampled white clover accessions are described in Kuo *et al.* (2024). In brief, the presence/absence of cyanogenic glucosides and linamarase were assessed by Feigl-Anger cyanogenesis assays following previously described protocols (Olsen *et al.*, 2007). For plants lacking given precursor, PAVs were confirmed by PCR-genotyping of *CYP79D15* (the first gene in the *Ac* gene cluster) and *Li* (Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen & Small, 2018). To test whether epistatic selection on the *Ac/ac* and *Li/li* polymorphisms affects genotype frequencies in nature (a prediction given that cyanogenesis requires the co-occurrence of dominant alleles at both loci) (Ennos, 1982; Kooyers & Olsen, 2012), we applied a Pearson's Chi-square test to compare observed frequencies of the two-locus genotypes with frequencies expected for unlinked genetic polymorphisms. Plants producing one or both cyanogenic precursors underwent further examination via genomic DNA qPCR to estimate CNVs (see Kuo *et al.*, 2024 for methodological details). The mean copy number of the *Ac* or *Li* locus was calculated for each of the 43 sampling locations before conducting the association analysis.

## Scanning for genome-wide environmental association

To assess the capability of our GBS dataset to capture adaptive signals, the dataset was first used in a genome-wide association study (GWAS) to map the loci that control the cyanogenesis polymorphism and confirm that the mapped locations correspond to the known genomic locations of *Ac* and *Li* (Olsen et al. 2021; Kuo et al. 2024). The GWAS was conducted using PLINK2, with the first ten principal components serving as the covariates.

*GEA analysis*. To identify the SNPs significantly associated with local environmental variables within the context of population structure, we employed a latent factor mixed model (LFMM) to examine the relationship between genetic and the environmental datasets (Caye *et al.*, 2019). LFMM requires no prior population information, and covariates (i.e., population structure or relatedness matrix) are calculated at the individual level, aligning well with our dataset's lack of discrete population structure (see Results). We set the number of latent factors (K) in the LFMM at three, determined as the best estimate from the ADMIXTURE analysis. Additionally, results from different K values (1-3) were compared. Output p-values were then transformed into *q*-values for false discovery rate (FDR) calibration using the 'qvalue' package in R.

*Differentiation scan.* To identify the SNPs significantly differentiated in the context of admixed population structure, we utilized PCadapt v4.3.5 to calculate the distance of a SNP to the overall genetic background (Privé *et al.*, 2020). The K value for PCadapt (the number of principal components) was set at two, following Cattell’s graphical rule as suggested in the original paper. A conventional *p*-value threshold was set at 5×10-8 based on the distribution of *p*-values in a Q-Q plot (**Fig. S1**).

All annotated genes in our reference genome (Kuo *et al.*, 2024) located within 25 Kbp of significant SNPs in the GEA and PCadapt analyses were manually examined based on descriptions in the UniProt database. Genes related to light response, developmental processes, stress, or pathogen response were retained as potential candidate genes for local environmental adaptation. Genomic regions identified in the landscape genomic analyses were also compared to locations of previously-identified fitness QTL for local adaptation in white clover (Wright et al. 2022) to assess the extent of overlap.

# Results

## Geographical distributions of SNP markers and environmental variation

After filtering to remove low-quality individuals and SNPs, the final dataset comprised 415 accessions (**Fig. 1**) and 345,762 biallelic SNPs. Three levels of pruning to remove SNPs in LD for population structure analysis resulted in the removal of 104,391 (mild pruning), 156,662 (medium pruning), and 226,663 (strong pruning) SNPs, respectively. The different degrees of LD pruning did not alter population structure inferences (**Fig. S2**); therefore, the dataset subjected to mild LD pruning was selected for subsequent analyses. Across the genome, the majority of SNPs with significant LD were within 25 Kbp, decreasing to background levels beyond 200 Kbp (**Fig. 2A**); this rapid LD decay is consistent with white clover's obligately outcrossing mating system as a self-incompatible species.

In ADMIXTURE analyses with assumed ancestral population numbers ranging from K = 1 to 8, the lowest cross-validation error was observed at K = 3 (**Fig. S3**). Most of the 43 sampling locations (see **Fig. 1**) exhibited admixed compositions at all assumed K values greater than 2, with the relative contributions of inferred ancestral populations varying by latitude (**Fig. 3A; Fig. S4; Fig. S5).** The five southernmost locations (GFL, MLA, NLA, ATX, TGA) differed from the rest of the range in maintaining a single ancestral population composition up to K=6. In line with the ADMIXTURE results, pairwise FST values between populations were very low overall (median FST =0.034), with somewhat greater differentiation detected for the comparisons that included the five southernmost populations (GFL, MLA, NLA, ATX, TGA; median FST = 0.078) (**Fig. S6**)**.** In the AMOVA, significant genetic differentiation was detected between the 43 sampling locations; however, the percentage of total genetic variation explained was low (2.89%) (**Table S2; Fig. S7**). Instead, most variation was distributed within locations — either between individuals (25.55%) or within them as heterozygosity (71.56%). Together these population structure assessments indicate that the 43 sampling locations do not represent distinct genetic populations, and that there is substantial genetic admixture throughout the range. Subsequent analyses were therefore based on this overall admixed population structure.

To assess environmental divergence among the 43 sampling locations, we performed a PCA of environmental variables. PC1 (explaining 53.78% of the total environmental variance) was strongly correlated with latitude, except for sampling locations in Washington state (BWA) and nearby Vancouver, British Columbia (VBC) which formed a distinct cluster along PC2 (explaining 19.41% of the total environmental variance) (**Fig. S8**). This environmental divergence is likely attributable to the moderate oceanic climate of the Pacific Northwest compared to inland locations of similar latitude.

To complement the population structure and environmental divergence analyses, we utilized a genetic PCA to visualize the population structure in relation to environmental variation. **Figure 3B** illustrates individual accessions plotted in PC1 and PC2 space. Most samples from a given location clustered together in genetic PCA space, indicating high genetic similarity among individuals within locations (**Fig. S9**). Consistent with the ADMIXTURE results, samples from southernmost locations (GFL, MLA, NLA, ATX, TGA) constituted a distinct genetic cluster (top-left corner of **Fig. 3B**). Notably, in a pattern that parallels the PCA of environmental data, PC1 for the genetic data also exhibited high collinearity with latitude, although this axis accounted for only a small portion of the total genetic variance (2.62%) (**Fig. 3C**). Considering that environmental variables are correlated with the latitude of sampling locations (**Fig. S8B**), this pattern suggests that latitude-associated environmental gradients could play some role in shaping and maintaining white clover population structure in North America.

## Associations between genetic differentiation, geographical distance, and environmental divergence

If the geographical distribution of genotypes is shaped by spatial environmental heterogeneity, we would predict that environmental differences between locations would account for more genetic differentiation between locations than geographical distance alone. To test this hypothesis, we assessed pairwise FST between the 43 sampling locations as a function of geographical distance or environmental divergence. As depicted in **Fig. 4A**, genetic differentiation is significantly correlated with geographical distance between locations (Mantel *r* = 0.21, *p* = 0.023); this pattern indicates that there is genetic isolation by distance (IBD) across the sampled range. Notably, however, genetic differentiation exhibits stronger associations with temperature-related environmental variables than with geographical distance alone (**Fig. 4B; Fig. S10**). This result indicates a stronger signal of isolation by environment (IBE) than IBD. Among the temperature-related variables, *Growing Degree Days (> 5°C)*, a common measure of thermal accumulation in the growing season (see **Fig. 1** inset), is the most effective in explaining genetic differentiation (Mantel *r* = 0.78, *p* < 0.001) (**Fig. 4B; Fig. S10**). Importantly, the associations with temperature-related variables remained significant even after accounting for the effect of geographical distance (partial Mantel test, **Fig. S11**). Together these results indicate that environmental distance is a better predictor of genetic differentiation between the sampling locations than geographical distance, providing positive support for the existence of IBE and local adaptation in North American white clover.

## Cyanogenesis distributions and environmental associations

The white clover cyanogenesis polymorphism is determined through the presence/absence of cyanogenic glucosides (controlled by the *Ac/ac* genetic polymorphism) and their hydrolyzing enzyme linamarase (controlled by the unlinked *Li/li* polymorphism) (see Methods). Therefore, four different cyanogenesis phenotypes (or “cyanotypes”) occur in nature: cyanogenic plants (“AcLi”); and acyanogenic plants that lack either cyanogenic glucosides (“acLi”), linamarase (“Acli”) or both components (“acli”). The 415 sampled accessions collectively included a total of 134 AcLi (32.29%), 121 Acli (29.16%), 49 acLi (11.81%), and 111 acli (26.75%) cyanotypes. As illustrated in **Fig. 1**, the frequency of cyanogenic (AcLi) cyanotypes is highest in the southernmost sampled locations, consistent with previously reported latitudinal cyanogenesis clines in North America (Daday, 1958; Kooyers & Olsen, 2012; Innes *et al.*, 2022).

The cyanogenic phenotype (AcLi) requires the production of both cyanogenic glucosides and linamarase. This creates the possibility that epistatic selection could be operating in nature, favoring plants that are either AcLi (cyanogenic) or acli (acyanogenic and not bearing any energetic costs of producing cyanogenic precursors), and disfavoring Acli and acLi plants (which bear some energetic costs of cyanogenesis but without the benefit of the chemical defense) (Ennos 1982; Kooyers & Olsen 2012). To test for evidence of epistatic selection, we compared our observed cyanotype frequencies to null expectations for two unlinked, independently-assorting polymorphisms. Consistent with the pattern expected under epistatic selection, the observed cyanotype frequencies showed a significant excess of AcLi and acli cyanotypes and deficits of acLi and Acli cyanotypes (χ2 = 18.29, df = 1, p = 1.897×10-5) (**Fig. 5C**). Interestingly, this non-random association between the *Ac* and *Li* loci was evident not only from cyanotype frequencies, but also in the form of significantly elevated interchromosomal LD between SNPs linked to each cyanogenesis locus compared to background levels (Kolmogorov-Smirnov (KS) test, p < 0.001) (**Fig. 2B**). The locus-specific pattern of this interchromosomal LD rules out the possibility that the nonrandom association could be an artifact of population structure. Our inference of epistatic selection at the cyanogenesis loci is in line with some but not all previous studies in white clover (Ennos 1982; reviewed by Kooyers & Olsen 2012).

The repeated evolution of climate-associated cyanogenesis clines in white clover is a well-documented indicator of rapid environmental adaptation in this species (Daday, 1958; Kooyers & Olsen, 2012). In a recent study, we determined that the North American latitudinal cyanogenesis cline not only reflects population frequencies of cyanogenic genotypes (corresponding to PAVs at the *Ac* and *Li* loci), but that both *Ac* and *Li* alleles also show clinal variation in tandem gene copy number, with these CNVs contributing quantitatively to the strength of the cyanogenic phenotype (Kuo *et al.*, 2024). Based on this knowledge of CNV clines, we sought to determine which of the IBE-associated environmental variables identified above could best explain *Ac* and *Li* distributions across the sampled locations. Taking into account both PAVs and CNVs, the cyanogenesis clines are best explained by *Growing Degree Days (> 5°C)* (**Fig. 5A,B**), with other temperature-related variables following closely (**Fig. S12**). In contrast, precipitation-related variables (which have previously been reported to show associations with cyanogenesis clines; Kooyers & Olsen 2013; Kooyers et al. 2014), exhibited considerably lower associations. Notably, *Growing Degree Days (> 5°C)* is the same environmental variable that best explains genome-wide IBE (**Fig. S10; Fig. S11**); this suggests a very close association between the cyanogenesis cline distribution and overall patterns of climatic adaptation in white clover.

## Genome-wide scans for selection signals

The ability to pinpoint the genetic locus underlying trait variation is a crucial prerequisite for association analyses. To establish that our SNP dataset could be successfully applied in environmental association mapping, we first used the cyanogenesis phenotype data from our samples to map the underlying *Ac* and *Li* loci. Consistent with their known genomic locations (Olsen et al. 2021; Kuo et al. 2024), the *Ac* locus was successfully mapped to Chr. 2 at 8.5-9.0 Mbp, and the *Li* locus was mapped to Chr. 12 at 28 Mbp (**Fig 5D,E**; **Fig. S13**). Additional peaks were detected in the homeologous positions of each cyanogenesis gene, which we attribute to highly similar homeologous sequences in these subgenome regions (see details in Kuo et al. 2024).

Having established the feasibility of association mapping with our dataset, we next performed landscape genomic analyses to identify genomic regions associated with local environmental adaptation. We focused on *Growing Degree Days (> 5°C)* based on the IBE results above. In a GEA analysis controlling for population structure (K = 3 in LFMM, based on ADMIXTURE results), we identified 442 out of 345,762 SNPs (0.1278%) that exhibited significant associations (q < 0.05) (**Fig. 6B**). In the genome-wide differentiation scan to identify SNPs that significantly diverge in the context of population structure (PCadapt), we detected 213 out of 241,371 SNPs (0.08825%) as significant outliers (**Fig. 6C**). Genomic regions jointly identified in the two landscape genomic analyses are located at the bottom of Chr. 1, bottom of Chr. 7, middle of Chr. 13, and bottom of Chr. 15 (discussed below).

Since the GWAS for the cyanogenesis phenotype and the GEA shared the same underlying SNP dataset, we could directly compare them to test for signal overlaps at the cyanogenesis loci. Notably, neither *Ac* nor *Li* showed significant associations in the GEA, although SNPs linked to the *Ac* locus on Chr. 2 exhibited a detectable peak that marginally approached significance (**Fig. 6B**). This result potentially suggests that cyanogenesis variation is not a major determinant of local adaptation. However, it is important to recognize that the lack of statistical significance at the cyanogenesis loci could be partially attributable to the strong collinearity between the cyanogenesis cline and population structure (**Fig. 1; Fig. S14**); by controlling for population structure (setting K = 3 in LFMM), the GEA would have diminished power to detect these collinear adaptive loci (Lotterhos & Whitlock, 2015; Hoban *et al.*, 2016; Yoder & Tiffin, 2017). Therefore, to assess this possibility, we re-ran the GEA without any population structure control (i.e., using a simple linear model) and also at K = 1 and 2. Both the *Ac* and *Li* loci exhibited clear peaks in these re-analyses; however, neither locus reached the significance threshold whereas multiple other genomic regions were highly significant (**Fig. S15**). These results remained unchanged even after excluding of samples from Washington state (BWA) and Vancouver (VBC), where the environmentally distinct climate could potentially dampen signal detection (**Fig. S16**). Together these findings suggest that the cyanogenesis polymorphism, while spatially distributed in a pattern consistent with local adaptation, is not the primary determinant of local adaptation in white clover.

If outlier SNPs in landscape genomic scans reflect true signals of local adaptation, many of these SNPs would be expected to occur within causal genes or to be linked to them. To assess these possibilities, we manually investigated all protein-coding genes located within 25 Kbp of the outlier SNPs from the GEA and PCadapt output based on gene annotations in the reference genome (Kuo *et al.*, 2024). In the significant outlier intervals of the GEA and the genome-wide differentiation scan, there were 491 genes and 321 genes, respectively. After filtering for potential functions related to light response, developmental processes, stress, or pathogen response, 21 genes and 11 genes remained, respectively (**Fig. 6A**). Among these, 8 genes were identified in both analyses; these include several with functions related to flowering time and other developmental and stress response processes that are plausible candidates for local adaptation. Detailed gene information is provided in the supplementary information (**Table S3)**.

Finally, we compared the results of the GEA and the genomic differentiation scan to the previously identified QTLs for local adaptation in white clover, which were identified in reciprocal common garden experiments conducted in the northern, central and southern United States (Wright *et al.*, 2022). The fitness-related QTLs were categorized into vegetative growth traits (leaf area), survival (mortality) and reproductive output (flowering duration, floral count). We found that a subset of the regions identified through the landscape genomic analyses overlap these previously-identified fitness QTLs (**Fig. 6**). Among the overlapping regions, some contain candidate genes that could be playing functional roles in local adaptation. For example, a region at the bottom of Chr. 15 overlaps a major QTL for reproductive fitness traits and also contains candidate genes for flowering time and other developmental processes that were identified in both the GEA and differentiation scan analysis (e.g., *MYB13*, *FER*; see **Table S3**).

Notably, the three analyses presented in **Fig. 6** all have the shared goal of searching for adaptive genetic variation, but are based on different underlying hypotheses and datasets. Candidate loci at the intersection of the three analyses may therefore be of particular interest for characterizing the molecular basis of local adaptation in white clover. Loci identified in only one of the analyses could still be important contributors to local adaptation. By comparison, the *Ac* and *Li* cyanogenesis loci, which are not present in any of these genomic regions of interest, appear to play a secondary role at best in determining local adaptation in the sampled white clover populations.

# Discussion

Repeatedly-evolved clines across environmental gradients are considered a hallmark of local adaptation. However, the relative importance of conspicuous clinally-varying phenotypes, compared to other potential contributors to local adaptation, is seldom evaluated. The present study was undertaken with the goal of bridging this gap, utilizing the well-studied cyanogenesis polymorphism in white clover, which has repeatedly evolved climate-associated cyanogenesis clines worldwide. In our analyses of wild samples covering much of the North American species range, we found clear evidence for local environmental adaptation. Genome-wide differentiation across the sampled locations is better explained by temperature-related climatic variables than by geographic distance alone, with *Growing Degree Days (> 5°C)* identified as the most closely correlated environmental variable (**Fig. 4A, B**; **Fig. S10**). This same climatic variable is also strongly correlated with cyanogenesis clinal variation (**Fig. S12**). However, landscape genomic analyses indicate only a negligible contribution for cyanogenesis variation and its underlying genes in local adaptation; instead, multiple other genomic regions are implicated (**Fig. 6**). This finding aligns with our earlier results from fitness QTL mapping and common garden experiments (Wright *et al.*, 2018; Wright *et al.*, 2022), where we also found no significant support for cyanogenesis variation as a determinant of local adaptation. Below we explore these results further, considering them in context of the white clover system and their broader implications for understanding the genetic basis of local adaptation.

## Genomic insights on local adaptation in North American white clover

White clover, widely grown as a forage and rotation crop in temperate climates because of its remarkable nitrogen fixation capabilities, was introduced worldwide from its native range in Europe beginning in the 17th century; it is now a common cultivated and naturalized species throughout much of the world (Zeven, 1991; Kjærgaard, 2003; Taylor, 2008; Santangelo *et al.*, 2022). In North America, where its presence dates to European colonization (and its abundance was already noted by the mid-18th century; Andrae & Hancock, 2016), studies of wild populations have revealed clear evidence of local adaptation following its introduction; these include multiple studies documenting latitudinal and elevational cyanogenesis clines that are correlated with temperature gradients (Ganders, 1990; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013), as well as studies that have directly measured fitness traits in common garden experiments (Wright *et al.*, 2018; Wright *et al.*, 2022). The present study, by providing a complementary landscape genomics perspective, allows for several additional insights into the factors shaping the population structure and local adaptation in these North American populations.

*Genetic admixture across North American populations.* First, with respect to population structure, North American white clover shows minimal evidence of geographical structuring into discrete genetic subpopulations. Instead, genetic admixture predominates across the sampled locations (**Fig. 3A**), with the vast majority of the total genetic variation (>97%) distributed within sampling locations rather than between them (**Table S2**). Consistent with these patterns, the level of genetic differentiation among sampling locations is very low overall (median pairwise FST = 0.034). Together these observations suggest that barriers to gene flow are low across the sampled range, but with some limitation to long-range dispersal. This inference is consistent with white clover's obligately-outcrossing mating system and its abundance in lawns and other mowed and grazed areas, which creates a near-continuous distribution across much of the continent (U.S. Department of Agriculture, 2024).

The one region where we detect some evidence of non-admixed population substructure is in the southernmost sampling locations (Texas, Louisiana, Florida and southern Georgia). These samples form a homogeneous genetic subpopulation at the optimal K=3 in ADMIXTURE analyses and at higher assumed K values (**Fig. 3A**; **Fig.** **S4**); they also form a distinct cluster in the PCA (although on an axis that explains < 3% of the total genetic variance) (**Fig. 3B**), and they collectively show higher levels of genetic differentiation from other samples in FST analyses (**Fig. S6**). The slight but detectable genetic distinctness of these southernmost samples could reflect the remnants of a genetically distinct ancestral subpopulation in the native species range that contributed differentially to the establishment southern US populations. Comparative population structure analyses with samples from native European range could be useful to test this hypothesis.

*Genetic differentiation reflects Isolation-by-environment (IBE).* While genome-wide SNPs reveal a clear pattern of geographical isolation-by-distance (IBD) across the sampled locations (**Fig. 4A**), genetic differentiation is better explained as a function of temperature-related climatic variables than geography alone (**Fig. 4B**; **Fig. S10**). This suggests that latitudinally related climatic variation acts as stronger barrier to the movement of some alleles across the landscape than geographical distance. Systems characterized by extensive gene flow across a large-scale, continuous environmental gradient provide ideal conditions for the emergence of local adaptation (Wang & Bradburd, 2014). This process can in principle involve either new mutations arising *in situ*, with limited gene flow between climatically distinct locations (Orsini *et al.*, 2013), or selection acting on standing genetic variation that then becomes sorted geographically (Fay & Wu, 2000). As an introduced species with a history of <500 years in North America, the latter mechanism is likely most important for white clover. White clover populations worldwide are characterized by high genetic diversity, reflecting its history of intentional, repeated introductions globally as a nitrogen-fixing forage and lawn plant (Santangelo *et al.*, 2022; Caizergues *et al.*, 2024). This demographic history would facilitate the introduction of standing variants which could then become sorted by local environment. Consistent with this selective scenario, our previous analyses of the *Ac/ac* and *Li/li* cyanogenesis polymorphisms have shown that cyanogenesis clines in introduced world regions have arisen through selection on pre-existing alleles that are found in the native species range (Olsen *et al.*, 2013; Kooyers & Olsen, 2014).

The specific climatic variable that we find best describes IBE in white clover, *Growing Degree Days (> 5°C)*, is a common measure of the thermal accumulation in the growing season. Since this variable provides data on an annual time scale, we further calculated a monthly measure of Growing Degree Days (> 5°C) in order to provide greater resolution for the active growing season (**Fig. S17**). This revealed that the cumulative temperature in spring is drastically different between latitudinal locations, which strongly suggests that phenological response to local seasonal cues likely plays a critical role in determining locally-adaptive fitness in white clover. Notably, this finding mirrors our previous inferences from fitness data collected in common garden experiments in the northern, central and southern US (Wright *et al.*, 2018; Wright *et al.*, 2022), which also point to phenological timing as a key factor in local adaptation (discussed below).

*Insights on cyanogenesis cline evolution.* Climate-associated cyanogenesis clines have been extensively documented and studied in native and introduced populations of white clover for more than 70 years (Daday, 1954b; Daday, 1954a; Daday, 1958; Ennos, 1982; Ganders, 1990; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013). This body of work has generated a wealth of information on the evolution and ecology of this adaptive chemical defense polymorphism. While the findings of the present study do not implicate the cyanogenesis polymorphism as a key factor in local climatic adaptation in white clover, they do nonetheless provide two important new insights into the cyanogenesis polymorphism and cline evolution.

First, taking into account the full range of molecular variation that underlies the cyanogenesis polymorphism (both the PAVs that underlie the classic *Ac/ac* and *Li/li* polymorphisms, and the tandem CNVs present within ‘dominant’ alleles at each locus; **Table S1**; Kuo et al. 2024), we find that the geographical distribution of cyanogenesis clines is best described by the same environmental variable that best describes genome-wide differentiation, i.e., *Growing Degree Days (>5°C)* (**Fig. S10: Fig. S12**). Previous studies of cyanogenesis clines have generally reported correlations with measures related to winter temperature (e.g., minimum winter temperature [MWT]; mean temperature of the coldest quarter [MTCQ]) (e.g., Kooyers & Olsen, 2012; Kooyers & Olsen, 2013; Innes *et al.*, 2022; Kuo *et al.*, 2024), a period of dormancy and potential cold-related winter mortality. Our findings here suggest that rather than winter cold stress, it is instead selective factors related to the active growing season that underlie cyanogenesis cline evolution. *Growing Degree Days* (> 5°C) is a common estimate of insect pest life cycle (Bonhomme, 2000; Cayton *et al.*, 2015); thus, it is possible that this variable captures the level of herbivore pressure in a local environment that would favor the cyanogenesis chemical defense — especially in the early growing season, when young shoots are most vulnerable (Wright *et al.*, 2018). Our inference that cyanogenesis clines reflect selection during the warmer growing season rather than winter cold stress is also in line with recent work that has found no support for an older ‘autotoxicity’ hypothesis, which had proposed that freezing-induced cyanide toxicity might differentially select against cyanogenic genotypes in colder climates (Daday, 1965; Kooyers *et al.*, 2018; Kuo *et al.*, 2023)

A second inference on the cyanogenesis polymorphism is that we find clear evidence for epistatic selection as a factor shaping the cyanogenesis polymorphism in natural populations. Whereas the cyanogenic phenotype requires the presence of both cyanogenic glucosides and linamarase within a plant, neither of the individual components is known to serve any function other than in cyanogenesis ((although cyanogenic glucosides could function as N storage compounds; see Gleadow & Moller, 2014); at the same time, production of both cyanogenic components is likely energetically costly (Daday, 1965; Kakes, 1989; Kooyers *et al.*, 2018). This creates selective scenario whereby plants that produce only one of the two components could be maladaptive, since they carry some of the energetic costs of cyanogenesis but without any benefits of the chemical defense. Previous studies of white clover populations in North America and elsewhere have found mixed evidence for epistatic selection (Ennos, 1982) (reviewed by Kooyers & Olsen, 2012). Here, with a large, representative sample of North American genotypes (N=415), we observe significant deficits in frequencies of Acli and acLi plants, and a corresponding excess of AcLi and acli plants (**Fig. 5C**), consistent with a model of balancing epistatic selection that favors either cyanogenic plants or those that lack both components. Unlike previous studies, which relied solely on cyanotype frequencies to test for epistatic selection, we have further documented LD between the *Ac* and *Li* loci using genome-wide SNP data — which reveal significant interchromosomal LD specifically for the neutral SNPs that are linked to *Ac* and *Li* (**Fig. 2B**). These data not only provide compelling evidence for a model of epistatic balancing selection at the cyanogenesis loci (and rule out any possibility that the nonrandom association between *Ac* and *Li* could be an artifact of population structure), but they also add white clover to a growing list of species with genomic evidence of interchromosomal LD resulting from epistatic selection (e.g., Hench *et al.*, 2019; Gupta *et al.*, 2023).

*Landscape genomic insights on local adaptation in white clover.* Our GEA and genome-wide differentiation scans do not indicate that selection on the cyanogenesis polymorphism plays a primary role in local climatic adaptation. While *Ac*-linked SNPs did show a detectable peak below the significance threshold in the GEA , there was no visible signal at *Li* locus **(Fig. 6B)**; moreover, neither cyanogenesis locus reached significance even after removing corrections for population structure, which can undermine the power of detection in a GEA (Lotterhos & Whitlock, 2015; Hoban *et al.*, 2016; Yoder & Tiffin, 2017) (**Fig. S15**). This lack of evidence that cyanogenesis plays a key role in local climatic adaptation is consistent with our previous inferences based on common garden fitness measurements using wild ecotypes and F2 mapping populations (Wright *et al.*, 2018; Wright *et al.*, 2022).

While not implicating cyanogenesis as critical for local adaptation, the landscape genomic analyses do provide valuable insights into the genomic mechanisms of local adaptation in this climatically widespread species. Based on the results (Fig. 6), it is likely that a relatively few loci are driving this process. For species such as white clover, characterized by outcrossing mating systems, large population sizes and extensive gene flow, our findings are consistent with the expectation that local adaptation evolves though selection on a relatively few genes of major effect (Yeaman *et al.*, 2023). In line with this theoretical prediction, we observed that less than one-half of 1% of the total SNPs in the GEA scan (0.1278%) and less than one-tenth of 1% of SNPs in the PCadapt analysis (0.08825%) exhibited significant signatures of local adaptation.

As for the identity of the targets of selection, we hypothesize that selection for local climatic adaptation has acted on genes with functions related to the phenological response to local seasonal cues during the active growing season. Candidate genes within the significantly associated genomic regions include, for example, the photoperiod regulatory genes ***FRS5*** (Lin & Wang, 2004; Li *et al.*, 2011; Ma & Li, 2018) and ***CKL10*** (Tan *et al.*, 2013), as well the temperature sensing gene ***VRN1***, which functions in vernalization response (Sheldon *et al.*, 2000; Levy *et al.*, 2002). Additionally, the candidates include genes that function upstream of the flowering time regulator ***FLC***, including ***PDP3*** (Zhou *et al.*, 2018) and ***FER*** (Wang *et al.*, 2020) (**Table S3**). Consistent with this hypothesis, previous results from our common garden and fitness QTL mapping studies indicated that local climatic adaptation across the US is closely tied to loci controlling the relative timing of vegetative growth *vs*. flowering over successive seasons (Wright *et al.*, 2018; Wright *et al.*, 2022). Common garden observations indicated that southern US populations experience favorable growing conditions very early in the spring, followed by a period of dormancy or mortality during the hot summer months; northern populations, in contrast, experience a later initiation of favorable growth conditions but are able to continue vegetative growth and flowering later into the summer. These climatic differences appear to favor genotypes with early investment in flowering in southern climates, and those with later flowering (after an extended period of vegetative growth) further north (Wright *et al.*, 2022). Similar to these observations for white clover, research in diverse other plant species has also pointed to connections between latitudinal or elevational temperature gradients and phenology (most notably flowering time) in the process of local adaptation, which can potentially lead to population structure and even the emergence of reproductive isolating barriers among adaptively diverged populations (Stinchcombe *et al.*, 2004; De Frenne *et al.*, 2013; Toftegaard *et al.*, 2015; Bucher *et al.*, 2018).

While our study provides important insights into the genomic basis of local adaptation in white clover, some limitations should be noted. First, GEAs, as association-based methods, can yield excessive false-positive signals (Booker *et al.*, 2024). Direct fitness measurement and/or functional validation are therefore required to further validate the findings of this study. In addition, the genotyping-by-sequencing (GBS) protocol, while cost-effective, is a non-randomized and reduced representation whole-genome re-sequencing approach. Although it provides an efficient method for obtaining genotypic information in non-model organisms, it compromises resolution for detecting causal variations in coding or regulatory sequences and is prone to high missing rates in variant callings among samples. As population-level pangenome analyses become cost-effective in the future, clearer insights on the causal basis of local adaptation in this species will be possible. Similarly, expanded phenotypic analyses beyond cyanogenesis (the sole phenotype examined in the present study) would substantially enhance the depth and resolution of our understanding of the key traits involved, such as growth rate, flowering time, and stress responses. Future investigations in white clover and other widespread plant species will benefit from the incorporation of both genome sequencing and comprehensive phenotyping approaches.

## Broader implications

Phenotypic clines across environmental gradients have long captured the attention of evolutionary biologists (Haldane, 1948), and clines that evolve repeatedly within a single species range have been considered a gold standard for illustrating parallel adaptive evolution (e.g., Fabian *et al.*, 2015; Thompson *et al.*, 2016; Gould & Stinchcombe, 2017; Lasne *et al.*, 2018; McGoey *et al.*, 2020). When supported by fitness measurements and genomic signatures of selection, such systems can provide compelling evidence of local environmental adaptation in response to spatial environmental heterogeneity. However, there is a critical distinction between demonstrating that a trait is evolving in response to natural selection, and demonstrating that this trait is most important for determining fitness in the local environment. Our findings here provide an important lesson in this regard: the white clover cyanogenesis clines may be charismatic, ecologically interesting and easily detected, but none of those factors means that this polymorphism is most important for determining local environmental adaptation in this widely-adapted species. It is only through unbiased assessments of the genomic and phenotypic contributors to local adaptation that the key contributors can be reliably identified.

# Acknowledgements

We thank the high school Biology teachers, plant scientists and others who participated in the crowd-sourced collection of wild white clover accessions across North America, including Miranda Horton, Sunita Crittenden, Felix Fritschi, Briana Gross, Nic Kooyers, Lisa Limeri, Mary Lyman-Onkka, Ed McAssey, John McDonald, John McIntire, Maria Monteros, Jim Mullen, Patricio Munoz, Sandra Branham, Anne Puzzo, Nicole Riddle, Trisha Spanbauer, Daniel Tabor, Rebecca Thomson Putnam, Kathryn Turner, Cynthia Vigueira, Kate Waselkov, Marshall Wedger, and Dan Cui Zhou. We especially thank Alexander Mahmoud (undergraduate student at Washington University in St. Louis) for processing wild accessions and maintaining them in greenhouse. We also thank Mike Dyer and other staff of the Washington University in St. Louis greenhouse facility for providing plant care. Finally, we thank the members of the Olsen lab group who provided valuable comments on earlier versions of the manuscript.

# Conflict of Interest

The authors declare no conflicts of interest.

# Author Contribution

Wen-Hsi Kuo and Limei Zhong analyzed the data, interpreted the results, and wrote the manuscript. Wen-Hsi Kuo and Limei Zhong contributed equally to this work. Sara J. Wright designed the experiments, collected and maintained the wild accessions in greenhouse, phenotyped the cyanogenesis trait, and prepared and sequenced the GBS libraries. David M. Goad helped with the bioinformatic analyses and provided suggestions at the early stage of the project. Kenneth M. Olsen conceived and designed the project, interpreted the results, and edited the manuscript. All authors read and approved the final manuscript.

# Reference

**Ågren J, Oakley CG, McKay JK, Lovell JT, Schemske DW. 2013.** Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**(52): 21077-21082.

**Alexander DH, Novembre J, Lange K. 2009.** Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**(9): 1655-1664.

**Andrae J, Hancock D. 2016.** University of Georgia. College of A, Environmental S, eds. White clover establishment and management guide: University of Georgia.

**Antoniazza S, Burri R, Fumagalli L, Goudet J, Roulin A. 2010.** Local adaptation maintains clinal variation in melanin-based coloration of european barn owls (*Tyto alba*). *Evolution* **64**(7): 1944-1954.

**Battlay P, Wilson J, Bieker VC, Lee C, Prapas D, Petersen B, Craig S, van Boheemen L, Scalone R, de Silva NP, et al. 2023.** Large haploblocks underlie rapid adaptation in the invasive weed *Ambrosia artemisiifolia*. *Nature Communications* **14**(1): 1717.

**Bonhomme R. 2000.** Bases and limits to using ‘degree.day’ units. *European Journal of Agronomy* **13**(1): 1-10.

**Booker TR, Yeaman S, Whiting JR, Whitlock MC. 2024.** The WZA: A window-based method for characterizing genotype-environment associations. *Molecular Ecology Resources* **24**(2): e13768.

**Browning BL, Zhou Y, Browning SR. 2018.** A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics* **103**(3): 338-348.

**Bucher SF, Konig P, Menzel A, Migliavacca M, Ewald J, Romermann C. 2018.** Traits and climate are associated with first flowering day in herbaceous species along elevational gradients. *Ecology and Evolution* **8**(2): 1147-1158.

**Caizergues AE, Santangelo JS, Ness RW, Angeoletto F, Anstett DN, Anstett J, Baena-Diaz F, Carlen EJ, Chaves JA, Comerford MS, et al. 2024.** Does urbanisation lead to parallel demographic shifts across the world in a cosmopolitan plant? *Molecular Ecology* **33**(7): e17311.

**Caye K, Jumentier B, Lepeule J, Francois O. 2019.** LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular biology and evolution* **36**(4): 852-860.

**Cayton HL, Haddad NM, Gross K, Diamond SE, Ries L. 2015.** Do growing degree days predict phenology across butterfly species? *Ecology* **96**(6): 1473-1479.

**Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015.** Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**: 7.

**Daday H. 1954a.** Gene frequencies in wild populations of *Trifolium repens* L. Distribution by latitude. *Heredity* **8**(1): 61-78.

**Daday H. 1954b.** Gene frequencies in wild populations of *Trifolium repens*. II. Distribution by altitude. *Heredity* **8**(3): 377-384.

**Daday H. 1958.** Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity* **12**(2): 169-184.

**Daday H. 1965.** Gene frequencies in wild populations of *Trifolium repens* L. IV. Mechanism of natural selection. *Heredity* **20**(3): 355-365.

**De Frenne P, Graae BJ, Rodríguez‐Sánchez F, Kolb A, Chabrerie O, Decocq G, De Kort H, De Schrijver A, Diekmann M, Eriksson O, et al. 2013.** Latitudinal gradients as natural laboratories to infer species' responses to temperature. *Journal of Ecology* **101**(3): 784-795.

**Dixon P. 2003.** VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**(6): 927-930.

**Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011.** A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**(5): e19379.

**Ennos RA. 1982.** Association of the cyanogenic loci in white clover. *Genetics Research* **40**: 65-72.

**Fabian DK, Lack JB, Mathur V, Schlotterer C, Schmidt PS, Pool JE, Flatt T. 2015.** Spatially varying selection shapes life history clines among populations of Drosophila melanogaster from sub-Saharan Africa. *Journal of Evolutionary Biology* **28**(4): 826-840.

**Fay JC, Wu C-I. 2000.** Hitchhiking under positive darwinian selection. *Genetics* **155**(3): 1405-1413.

**Fick SE, Hijmans RJ. 2017.** WorldClim 2: new 1‐km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12): 4302-4315.

**Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011.** A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**(6052): 86-89.

**Ganders FR. 1990.** Altitudinal clines for cyanogenesis in introduced populations of white clover near Vancouver, Canada. *Heredity* **64**(3): 387-390.

**Gefaell J, Vigo R, González-Vázquez AH, Galindo J, Rolán-Alvarez E. 2023.** Temporal stability and directional change in a color cline of a marine snail from NW Spain. *Current Zoology*.

**Germino MJ, Moser AM, Sands AR. 2019.** Adaptive variation, including local adaptation, requires decades to become evident in common gardens. *Ecological Applications* **29**(2): e01842.

**Gleadow RM, Moller BL. 2014.** Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annual review of plant biology* **65**: 155-185.

**Gould BA, Stinchcombe JR. 2017.** Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. *Molecular Ecology* **26**(1): 92-106.

**Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019.** Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

**Grillo MA, Li C, Hammond M, Wang L, Schemske DW. 2013.** Genetic architecture of flowering time differentiation between locally adapted populations of *Arabidopsis thaliana*. *New Phytologist* **197**(4): 1321-1331.

**Guerrero J, Andrello M, Burgarella C, Manel S. 2018.** Soil environment is a key driver of adaptation in *Medicago truncatula:* new insights from landscape genomics. *New Phytologist* **219**(1): 378-390.

**Gugger PF, Liang CT, Sork VL, Hodgskiss P, Wright JW. 2018.** Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications* **11**(2): 231-242.

**Gupta S, Harkess A, Soble A, Van Etten M, Leebens‐Mack J, Baucom RS. 2023.** Interchromosomal linkage disequilibrium and linked fitness cost loci associated with selection for herbicide resistance. *New Phytologist*.

**Haldane JBS. 1948.** The theory of a cline. *Journal of Genetics* **48**(3): 277-284.

**Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011.** Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**(6052): 83-86.

**Hench K, Vargas M, Hoppner MP, McMillan WO, Puebla O. 2019.** Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology and Evolution* **3**(4): 657-667.

**Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016.** Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist* **188**(4): 379-397.

**Hof AEvt, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016.** The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**(7605): 102-105.

**Hughes MA. 1991.** The cyanogenic polymorphism in *Trifolium repens* L. (white clover). *Heredity* **66**(1): 105-115.

**Innes SG, Santangelo JS, Kooyers NJ, Olsen KM, Johnson MTJ. 2022.** Evolution in response to climate in the native and introduced ranges of a globally distributed plant. *Evolution* **76**(7): 1495-1511.

**Kakes P. 1989.** An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. *Theoretical and Applied Genetics* **77**(1): 111-118.

**Kamvar ZN, Tabima JF, Grünwald NJ. 2014.** Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**: e281.

**Kawecki TJ, Ebert D. 2004.** Conceptual issues in local adaptation. *Ecology Letters* **7**(12): 1225-1241.

**Kjærgaard T. 2003.** A plant that changed the world: The rise and fall of clover 1000-2000. *Landscape Research* **28**(1): 41-49.

**Kooyers NJ, Gage LR, Al-Lozi A, Olsen KM. 2014.** Aridity shapes cyanogenesis cline evolution in white clover (*Trifolium repens* L.). *Molecular Ecology* **23**(5): 1053-1070.

**Kooyers NJ, Hartman Bakken B, Ungerer MC, Olsen KM. 2018.** Freeze-induced cyanide toxicity does not maintain the cyanogenesis polymorphism in white clover (*Trifolium repens*). *American Journal of Botany* **105**(7): 1224-1231.

**Kooyers NJ, Olsen KM. 2012.** Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Molecular Ecology* **21**(10): 2455-2468.

**Kooyers NJ, Olsen KM. 2013.** Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.). *Heredity* **111**(6): 495-504.

**Kooyers NJ, Olsen KM. 2014.** Adaptive cyanogenesis clines evolve recurrently through geographical sorting of existing gene deletions. *Journal of Evolutionary Biology* **27**(11): 2554-2558.

**Koski MH, Galloway LF. 2018.** Geographic variation in pollen color is associated with temperature stress. *New Phytologist* **218**(1): 370-379.

**Kujala ST, Knürr T, Kärkkäinen K, Neale DB, Sillanpää MJ, Savolainen O. 2017.** Genetic heterogeneity underlying variation in a locally adaptive clinal trait in *Pinus sylvestris* revealed by a Bayesian multipopulation analysis. *Heredity* **118**(5): 413-423.

**Kuo W-H, Small LL, Olsen KM. 2023.** Variable expression of cyanide detoxification and tolerance genes in cyanogenic and acyanogenic white clover (*Trifolium repens*). *American Journal of Botany* **110**(10): e16233.

**Kuo W-H, Wright SJ, Small LL, Olsen KM. 2024.** *De novo* genome assembly of white clover (*Trifolium repens* L.) reveals the role of copy number variation in rapid environmental adaptation. *Research Square*.

**Lascoux M, Glémin S, Savolainen O 2016.** Local adaptation in plants. *Encyclopedia of Life Sciences*,1-7.

**Lasky JR, Josephs EB, Morris GP. 2022.** Genotype–environment associations to reveal the molecular basis of environmental adaptation. *The Plant Cell* **35**(1): 125-138.

**Lasne C, Hangartner SB, Connallon T, Sgro CM. 2018.** Cross-sex genetic correlations and the evolution of sex-specific local adaptation: Insights from classical trait clines in *Drosophila melanogaster*. *Evolution* **72**(6): 1317-1327.

**Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. 2002.** Multiple Roles of Arabidopsis *VRN1* in Vernalization and Flowering Time Control. *Science* **297**(5579): 243-246.

**Li G, Siddiqui H, Teng Y, Lin R, Wan XY, Li J, Lau OS, Ouyang X, Dai M, Wan J, et al. 2011.** Coordinated transcriptional regulation underlying the circadian clock in Arabidopsis. *Nature Cell Biology* **13**(5): 616-622.

**Li H, Durbin R. 2009.** Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

**Lin R, Wang H. 2004.** Arabidopsis *FHY3*/*FAR1* gene family and distinct roles of its members in light control of Arabidopsis development. *Plant Physiology* **136**(4): 4010-4022.

**Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009.** On the origin and spread of an adaptive allele in deer mice. *Science* **325**(5944): 1095-1098.

**Lotterhos KE, Whitlock MC. 2015.** The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* **24**(5): 1031-1046.

**Ma L, Li G. 2018.** FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in Arabidopsis growth and development. *Frontiers in Plant Science* **9**.

**Martin M. 2011.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**(1): 3.

**McGoey BV, Hodgins KA, Stinchcombe JR. 2020.** Parallel flowering time clines in native and introduced ragweed populations are likely due to adaptation. *Ecology and Evolution* **10**(11): 4595-4608.

**Mullen LM, Hoekstra HE. 2008.** Natural selection along an environmental gradient: A classic cline in mouse pigmentation. *Evolution* **62**(7): 1555-1570.

**Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF. 2021.** Dual‐species origin of an adaptive chemical defense polymorphism. *New Phytologist* **232**(3): 1477-1487.

**Olsen KM, Hsu SC, Small LL. 2008.** Evidence on the molecular basis of the *Ac*/*ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**(1): 517-526.

**Olsen KM, Kooyers NJ, Small LL. 2013.** Recurrent gene deletions and the evolution of adaptive cyanogenesis polymorphisms in white clover (*Trifolium repens* L.). *Molecular Ecology* **22**(3): 724-738.

**Olsen KM, Small LL. 2018.** Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytologist* **219**(2): 757-766.

**Olsen KM, Sutherland BL, Small LL. 2007.** Molecular evolution of the *Li*/*li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology* **16**(19): 4180-4193.

**Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013.** Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**(24): 5983-5999.

**Pebesma EJ. 2018.** Simple features for R: standardized support for spatial vector data. *The R Journal* **10**(1): 439.

**Poplin R, Ruano-Rubio V, Depristo MA, Fennell TJ, Carneiro MO, Van Der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2017.** Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*.

**Price N, Moyers BT, Lopez L, Lasky JR, Monroe JG, Mullen JL, Oakley CG, Lin J, Ågren J, Schrider DR, et al. 2018.** Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **115**(19): 5028-5033.

**Privé F, Luu K, Vilhjálmsson BJ, Blum MGB. 2020.** Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular biology and evolution* **37**(7): 2153-2154.

**Santangelo JS, Ness RW, Cohan B, Fitzpatrick CR, Innes SG, Koch S, Miles LS, Munim S, Peres-Neto PR, Prashad C, et al. 2022.** Global urban environmental change drives adaptation in white clover. *Science* **375**(6586): 1275-1281.

**Savolainen O, Lascoux M, Merilä J. 2013.** Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**(11): 807-820.

**Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES. 2000.** The molecular basis of vernalization: The central role of *FLOWERING LOCUS C* (*FLC*). *Proceedings of the National Academy of Sciences of the United States of America* **97**(7): 3753-3758.

**Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004.** A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences of the United States of America* **101**(13): 4712-4717.

**Tan ST, Dai C, Liu HT, Xue HW. 2013.** Arabidopsis casein kinase1 proteins CK1.3 and CK1.4 phosphorylate cryptochrome2 to regulate blue light signaling. *The Plant Cell* **25**(7): 2618-2632.

**Taylor NL. 2008.** A century of clover breeding developments in the united states. *Crop Science* **48**(1): 1-13.

**Thompson KA, Renaudin M, Johnson MT. 2016.** Urbanization drives the evolution of parallel clines in plant populations. *Proceedings of the Royal Society B: Biological Sciences* **283**(1845).

**Title PO, Bemmels JB. 2018.** ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* **41**(2): 291-307.

**Toftegaard T, Posledovich D, Navarro‐Cano JA, Wiklund C, Gotthard K, Ehrlén J. 2015.** Variation in plant thermal reaction norms along a latitudinal gradient – more than adaptation to season length. *Oikos* **125**(5): 622-628.

**U.S. Department of Agriculture NRCS 2024**. *Trifolium repens* L.

**Wang IJ, Bradburd GS. 2014.** Isolation by environment. *Molecular Ecology* **23**(23): 5649-5662.

**Wang L, Yang T, Lin Q, Wang B, Li X, Luan S, Yu F. 2020.** Receptor kinase FERONIA regulates flowering time in Arabidopsis. *BMC Plant Biol* **20**(1): 26.

**Wang Y, Zhang L, Zhou Y, Ma W, Li M, Guo P, Feng L, Fu C. 2023.** Using landscape genomics to assess local adaptation and genomic vulnerability of a perennial herb *Tetrastigma hemsleyanum* (Vitaceae) in subtropical China. *Frontiers in Genetics* **14**: 1150704.

**Whitlock R, Hipperson H, Mannarelli M, Burke T. 2008.** A high-throughput protocol for extracting high-purity genomic DNA from plants and animals. *Molecular Ecology Resources* **8**(4): 736-741.

**Wittkopp PJ, Smith-Winberry G, Arnold LL, Thompson EM, Cooley AM, Yuan DC, Song Q, McAllister BF. 2011.** Local adaptation for body color in *Drosophila americana*. *Heredity* **106**(4): 592-602.

**Wright SJ, Cui Zhou D, Kuhle A, Olsen KM. 2018.** Continent-wide climatic variation drives local adaptation in North American white clover. *Journal of Heredity* **109**(1): 78-89.

**Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022.** Genetic trade‐offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

**Yeaman S, Whiting J, Booker T, Rougeux C, Lind B, Singh P, Lu M, Huang K, Whitlock M, Aitken S. 2023.** Core genes driving climate adaptation in plants. *Research Square*.

**Yoder JB, Tiffin P. 2017.** Effects of gene action, marker density, and timing of selection on the performance of landscape genomic scans of local adaptation. *Journal of Heredity* **109**(1): 16-28.

**Zeven AC. 1991.** Four hundred years of cultivation of Dutch white clover landraces. *Euphytica* **54**(1): 93-99.

**Zhou JX, Liu ZW, Li YQ, Li L, Wang B, Chen S, He XJ. 2018.** Arabidopsis PWWP domain proteins mediate H3K27 trimethylation on FLC and regulate flowering time. *Journal of Integrative Plant Biology* **60**(5): 362-368.

# Data Accessibility and Benefit-Sharing

## **Data Accessibility Statement**

Raw GBS sequence reads are deposited in NCBI SRA (BioProject: PRJNA1014483). The complied SNP dataset and LFMM result are deposited in Dryad (DOI: 10.5061/dryad.s7h44j1fd). The R scripts for generating the main figures are available at GitHub (https://github.com/kuowenhsi/clover\_script).

## **Benefit-Sharing Statement**

Benefits Generated: This study benefited from the collaboration of numerous high school teachers who contributed wild population samples (see Acknowledgements). To foster this partnership, the Olsen laboratory distributes "Clover Kits" to high schools across the United States through the Clover Project (<https://cloverlab.weebly.com>). Each kit contains essential tools and comprehensive instructions designed for high school educators and students to explore the biology of the cyanogenesis polymorphism in white clover, which is prevalent in grasslands, parks, and campuses nationwide. These kits provide opportunities for both teachers and students to engage firsthand with fundamental concepts of inheritance, population genetics, ecology, and plant biology.